COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 04-23	FOR NSF USE ONLY
NSF 05-628 02/02/06	NSF PROPOSAL NUMBER
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) SES - INNOVATION & ORG CHANGE (IOC)	

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
				002257350	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)
150532081	☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
Syracuse University	Syracuse University 113 Bowne Hall Syracuse, NY. 132441200
AWARDEE ORGANIZATION CODE (IF KNOWN) 0028829000	
NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE	ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE
PERFORMING ORGANIZATION CODE (IF KNOWN)	

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)	☐ SMALL BUSINESS ☐ FOR-PROFIT ORGANIZATION	☐ MINORITY BUSINESS ☐ WOMAN-OWNED BUSINESS	☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT **Group maintenance for cyber-infrastructure-supported distributed groups**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 299,958	24 months	07/01/06	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

- ☐ BEGINNING INVESTIGATOR (GPG I.A)
- ☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)
- ☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.1.d)
- ☐ HISTORIC PLACES (GPG II.C.2.j)
- ☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)
- ☐ VERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date ____________
- ☒ HUMAN SUBJECTS (GPG II.D.6) Exemption Subsection ____________ or IRB App. Date Pending
- ☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.j) ____________
- ☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.G.1)

PI/PD DEPARTMENT	PI/PD POSTAL ADDRESS
School of Information Studies	Syracuse University Information Studies 4-206 Centre for Science and Technology Syracuse, NY 132444100 United States
PI/PD FAX NUMBER 315-443-5806	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME Kevin Crowston	PhD	1991	315-443-1676	crowston@syr.edu
CO-PI/PD Derrick L Cogburn	PhD	1996	315-443-5441	dcogburn@syr.edu
CO-PI/PD Robert L Heckman	Ph.D.	1993	315-443-4479	rheckman@syr.edu
CO-PI/PD Elizabeth D Liddy	PhD	1988	315-443-5484	liddy@mailbox.syr.edu
CO-PI/PD				

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the individual applicant or the authorized official of the applicant institution is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 04-23. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

In addition, if the applicant institution employs more than fifty persons, the authorized official of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of Grant Policy Manual Section 510; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Appendix C of the Grant Proposal Guide.

Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency? Yes ☐ No ☒

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Appendix D of the Grant Proposal Guide.

Certification Regarding Lobbying

This certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, ''Disclosure of Lobbying Activities,'' in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME		

TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER

*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.

Project Summary: Group Maintenance in Cyber-infrastructure-supported Distributed Groups

We propose a study of the ways in which members of distributed groups create and maintain a social environment that enables and motivates members to work together using cyber-infrastructure [5, 12]. Distributed groups are networks of geographically dispersed individuals working together over time towards a common goal. Distributed work has a long history [e.g., 87], but recent advances in information and communication technologies (ICT) have been crucial enablers for extension of this organizational form [1]. Cyber-infrastructure is an emerging concept that refers to the constellation of ICT systems designed to support the communication, coordination, collaboration, data collection, storage, analysis and dissemination needs of distributed groups. As a result of these technical innovations, distributed groups are becoming more common in all kinds of organizations [82]. However, the distance between group members—geographic, organizational and social distance—challenges members to maintain the social relationships necessary for the group to be effective [91]. A substantial and growing knowledge base exists for understanding geographically-distributed collaboration in science and in the workplace [41, 118]. However, much less is known about the social aspects of teamwork in distributed groups. To fill this gap, the proposed research addresses the following general research question:

What kinds of group maintenance behavior enable members of cyber-infrastructure-supported distributed groups to work together most effectively, and how?

Expected intellectual merit

The intellectual merit of the proposed research is that it addresses a fundamental problem in organizational behavior, namely group maintenance, in a novel setting, namely distributed groups working together using cyber-infrastructure, to advance our understanding of the effects of interpersonal relationships on the functioning, effectiveness and innovation of groups who rely on innovative applications of ICT. To address our research question, we develop a conceptual model of group maintenance behavior and apply it to the study of functioning distributed groups in three related but distinct empirical settings, namely 1) scientific research collaboratories; 2) transnational policy networks; and 3) FLOSS (free/libre open source software) development groups. These types of groups have been chosen because each involves collaborations between geographically and organizationally separated members, carried out primarily via cyber-infrastructure, in order to accomplish shared tasks that produce some kind of innovation. A novel aspect of our proposal is the application of Natural Language Processing (NLP) approaches to facilitate the qualitative social-science analysis of large-scale digital data to assess group maintenance behavior.

Expected broader impacts

The proposed research will have broader impacts of several types. First, the project will <u>benefit society</u> by providing generalizable knowledge to improve the effectiveness of distributed groups. Distributed groups are increasingly common in a variety of settings, including multi-disciplinary university centers, industrial research departments, and civil-society and non-governmental organizations, and are used for a variety of tasks, including research, development and engineering. As a result, the project results should be particularly pertinent to the effective organization and management of scientific efforts that involve shared technological resources, particularly cyber-infrastructure resources. To ensure that our study has a significant impact, we plan to <u>broadly disseminate</u> results through journal publications, conferences, workshops and on our Web pages, as well as through our interaction with the leaders and members of distributed teams. Findings from the study might also be used to enhance the way computer-mediated communication technologies (CMC) are used to support distance education or scientific collaboration, which are emerging applications of distributed teams. Finally, the project will <u>promote teaching, training, and learning</u> by providing an opportunity for students to work on research teams, utilize their competencies and develop new skills in data collection, model development and data analysis.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

Group maintenance in cyber-infrastructure-supported distributed groups

We propose a study of the ways in which members of distributed groups create and maintain a social environment that enables and motivates members to work together using cyber-infrastructure [5, 12]. Distributed groups are networks of geographically dispersed individuals working together over time towards a common goal. Distributed work has a long history [e.g., 87], but recent advances in information and communication technologies (ICT)—from email, instant messaging and presence awareness systems, to web conferencing and easy-to-use content management systems—have been crucial enablers for development of this organizational form [1]. Cyber-infrastructure is an emerging concept that refers to the constellation of ICT systems designed to support the communication, coordination, collaboration, data collection, storage, analysis and dissemination needs of distributed groups. As a result of these technical innovations, distributed groups are becoming more common in all kinds of organizations [82]. However, the distance between group members—geographic, organizational and social distance—challenges members to maintain the social relationships necessary for the group to be effective [91]. A substantial and growing knowledge base exists for understanding geographically-distributed collaboration in science and in the workplace [41, 118]. However, much less is known about the social aspects of teamwork in distributed groups and how they contribute to teams' performance. To fill this gap, the proposed research addresses the following general research question:

What kinds of group maintenance behavior enable members of cyber-infrastructure-supported distributed groups to work together most effectively, and how?

Expected intellectual merit

The intellectual merit of the proposed research is that it addresses a fundamental problem in organizational behavior, namely group maintenance, in a novel setting, namely distributed groups working together using cyber-infrastructure, to advance our understanding of the effects of interpersonal relationships on the functioning, effectiveness and innovation of groups who rely on innovative applications of ICT. To address our research question, we develop a conceptual model of group maintenance behavior and apply it to the study of functioning distributed groups in three related but distinct empirical settings. A novel aspect of our proposal is the application of Natural Language Processing (NLP) approaches to facilitate the social-science analysis of large-scale digital data to assess group maintenance behavior. The first two aspects of the study are briefly introduced here and expanded on in the following sections of the proposal.

Theoretical foundation. Though different research streams have used different labels, researchers have commonly differentiated between two broadly defined types of group behavior: *task-oriented* behavior and *relational* or *group maintenance* behavior. Group maintenance behavior is discretionary, pro-social, relation-building behavior that enables group members to more easily trust and cooperate with one another, based on the expectation of the future cooperation of others [98], what game theorists call the "shadow of the future" [7]. Such behavior is closely related to an array of prosocial behaviors that have been identified by organizational theorists in various contexts: consideration, expressive behavior, or relational behavior in leadership research [57, 120, 121]; social presence in community of inquiry literature [42, 100]; social-emotional behavior, face work, or social presence in computer-mediated communications (CMC) research [46, 84, 88]; and organizational citizenship behavior (OCB), relation-oriented behavior, supportiveness, conflict management in organizational research [44, 53, 94]. Whatever the label, group maintenance behavior is important because it is believed to be associated with a number of desirable group and organizational outcomes.

Empirical settings. The proposed study is set in the context of cyber-infrastructure-supported distributed groups, an increasingly common organizational form enabled by technological advances and driven by the need for collaboration within industry, local or international policy communities, and scientific communities. We focus in particular on distributed groups whose goal is creating a collective innovation, since this type of group requires intensive collaboration and decision making that can benefit from access to distributed knowledge and expertise. However, the distance between members and the limited opportunities for interaction provided by cyber-infrastructure suggest that many of the traditional tactics of group maintenance will be difficult to apply, even though "the social glue of good relations among participants" is still critical [11]. To develop generalizable findings, we will compare and contrast group maintenance in three different types of distributed groups that rely on cyber-infrastructure: 1) scientific research collaboratories; 2) transnational policy networks; and 3) free/libre open source software (FLOSS) development groups. The rationale for our choice of these types of distributed groups will be described in detail below in the section on study design.

Expected broader impacts

In addition to the expected intellectual contributions described above, the proposed research will benefit society by providing generalizable knowledge to improve the effectiveness of distributed groups. Distributed groups are increasingly common in a variety of settings, including multi-disciplinary university centers, industrial research departments, and civil-society and non-governmental organizations, and are used for a variety of tasks, including research, development and engineering (see, for example, the attached letters of support from the directors of research collaboratories expressing their interest in the proposed research). In addition, the proposed project will have an impact by promoting teaching, training, and learning by students involved in the research project (note that the majority of the requested funding supports students).

The remainder of this proposal is organized into four sections. In section 1, we develop a conceptual model for our study, drawing on various research literatures that address the phenomenon of group maintenance. In section 2, we present the study design, with details of the data collection and analysis plans, and describe how our research will integrate social science and natural language processing (NLP). In section 3, we present the project management plan. We conclude in section 4 by sketching the intellectual merits and expected broader impacts of our study and by reviewing results of prior NSF support.

1. Conceptual Development

In this section we develop the conceptual framework for our study, building on and adding to existing literature drawn from multiple disciplines. We are interested in the phenomenon of group maintenance, that is, discretionary, prosocial, relation-building behavior that enables group members to more easily trust and cooperate with one another. Voluntary groups, whether part of businesses, societal communities or research communities, will not last long if members are dissatisfied and ineffective collaborators. Groups that last over time develop a social environment that is conducive to accomplishing group tasks, and to the social needs of individual members. This social environment includes open communication among the group members, support of the group members' needs, an effective conflict resolution process and commitment by the group to minimize process losses [i.e, group synergy, as defined by 50]. The development of a supportive social environment is particularly problematic in distributed groups in which members have few opportunities to meet and work together face-to-face.

We define group maintenance behavior as discretionary, prosocial, relation-building behavior that is not explicitly task oriented, or even directly necessary for the completion of a group's task. While such behavior may be closely intertwined with task-oriented behavior, and while both functions may even be evident in a single act, it is possible to distinguish between group maintenance and task-oriented functions. Researchers have identified an array of discretionary, prosocial behaviors that contribute to the creation of an environment that supports a work group's task related activities. While different labels have been used to describe these behaviors, they share several characteristics that will be important to the research we propose here. We will examine these characteristics by reviewing several streams of literature that shed light on their commonalities. We first discuss research in leadership, group performance, and organizational citizenship behavior that helps us to understand the general nature of group maintenance behavior. We then turn to research that more specifically addresses group maintenance behavior that occurs via cyber-infrastructure, as in distributed groups.

Leadership. Most functionalist leadership theories make a broad distinction between *task leadership behaviors* and *group maintenance leadership* behaviors. The former are concerned with organizing, coordinating and performing the task(s) that constitute the group's primary work, while the latter are concerned with maintaining group morale, motivation and communication. Bales [9] believed that the functions of task and group maintenance behaviors are opposed, and that groups should strive to find a balance or equilibrium between them. The opposition between task and maintenance behaviors also suggested to Bales that it would be more likely that different people would emerge to perform task and maintenance roles [95]. In addition to the task and group maintenance functions that leadership must satisfy, Ancona and Caldwell [2] argued that there are also leadership functions involved with maintaining relations with individuals and groups outside the group. Finally, in a distributed group, where members make diverse knowledge contributions [48], it may be useful to distinguish between two types of task roles, *procedural* and *substantive*. *Procedural behaviors* are those involved in coordinating the group's work (e.g., scheduling, dividing labor, creating processes, choosing and maintaining collaboration infrastructure) while *substantive behaviors* are those that actually accomplish the group's work (e.g., idea generation, evaluation, integration, synthesis). Thus, leaders may exercise their influence by means of their substantive expertise as well as through their coordinating and directing activities. In summary, according to this body of research, social, or group maintenance behavior may be performed by different group members at different times, and can be expected to be distinguishable from, yet complement task leadership [123].

Group performance. Schutz identified three functions critical to the effectiveness of a group as a social system: the group's relations with other people and other groups; members' relations with one another; and members' interdependent work toward a shared goal. Integrating and building on Schutz [104] and Roby [99], Walton and Hackman [112] identified five main work-group functions: social, interpretive, regulative, agency and task management functions. The social function of a group is to humanize the workplace for its members, and the interpretive function is to create a social reality for members. The regulative function concerns the generation and enforcement of norms, the agency function is to influence other organizational entities, and the task management function is to coordinate resources to perform the tasks required to reach the group's goals. Thus the functions of the group can be summarized as coordinating efforts, resources and other entities to perform group tasks (regulative, agency and task management functions), while at the same time creating a social and humane work environment that meets members' social needs (social and interpretive functions.) When the group members' behaviors and structures satisfy the social needs of group members, they can create what Hackman [50] called group synergy. Hackman's model, shown in Figure 1, describes group synergy as a primary moderator that improves group effectiveness. A social environment that includes a high level of group synergy will facilitate individual contribution to tasks and goals. We note that group synergy is achieved through the contributions of a group's individual members. For example, Gladstein [44] identified supportiveness, conflict management, and open communication as processes that create group synergy, suggesting these as aspects of group maintenance behavior. Without individuals' willingness to contribute group maintenance behavior, synergy will be low and the group will not be effective.

Organizational citizenship behavior. Organizational citizenship behavior (OCB) has been defined as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and in the aggregate promotes the efficient and effective functioning of the organization" [94]. Several dimensions of OCB have been identified, including *helping* (behavior in which the immediate beneficiary is a specific individual person), *compliance* (general adherence to the spirit of the rules or norms that define a cooperative system), *sportsmanship* (putting up with minor grievances and inconveniences without complaining), *civic virtue* (responsible, constructive involvement in governance processes) and *courtesy* (avoiding practices that make other people's work harder) [47, 62, 93, 106, 116]. OCB has been widely studied in formal organizational settings for a number of years. This research suggests that OCB is closely related to positive attitudes such as job satisfaction. Theorists have also proposed that dispositional traits (i.e., personality) would predict OCB, but the bulk of the empirical research on this issue does not support this relationship

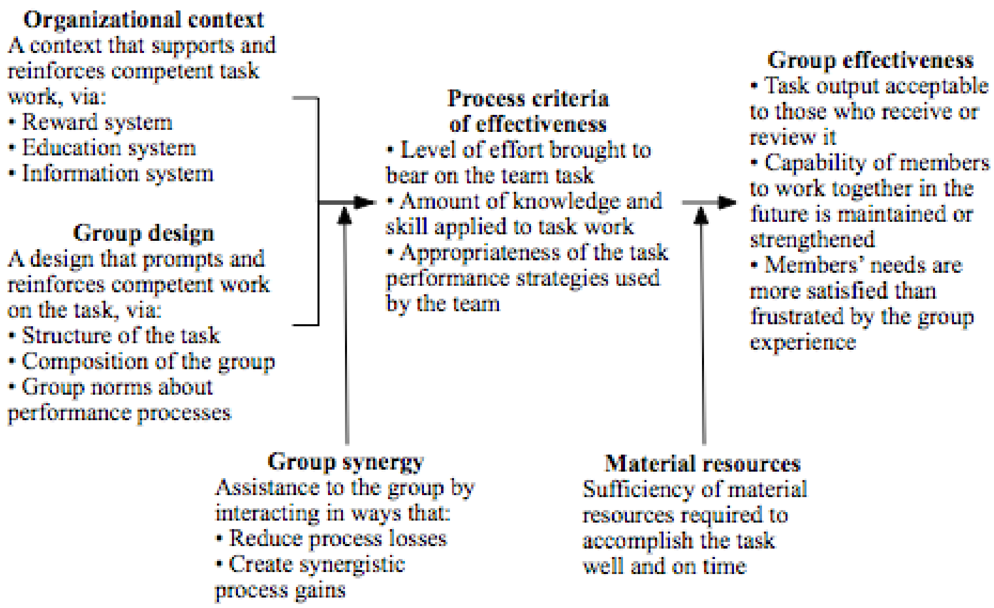


Figure 1. Hackman's [50] normative model of group effectiveness

[94]. Organizational citizenship behavior has also been associated positively with performance quantity and quality, financial efficiency, and good customer service [94]. Thus we find a large body of research that associates this form of discretionary prosocial organizational behavior with desirable group outcomes and characteristics. Because the majority of this research has been cross-sectional and correlational, theorists have been careful to point out that we cannot say with certainty whether variables such as job satisfaction are antecedents of OCB, outcomes of OCB, or together with OCB caused by a third variable. Nevertheless, evidence for a relationship between this form of group maintenance behavior and positive group characteristics and outcomes continues to grow.

In summary, research on leadership, group performance, and organizational citizenship behavior suggest that the performance of groups and organizations is strongly related to the maintenance of a positive, constructive social environment. As well, this research has identified a variety of group maintenance behavior that may contribute to the development and preservation of such an environment. However, in distributed groups, the opportunities for group maintenance behavior are limited by the fact that interactions are predominantly mediated by ICT. Heckman and Annabi [52] suggest that the lack of informal, face-to-face communication presents challenges for collaboration and learning in distributed groups. In the remainder of this section, therefore, we turn to research that has attempted to identify group maintenance behavior carried out via ICT. We first briefly review research on *virtual teams* before turn to research on computer-mediated asynchronous discourse, specifically, *community of inquiry* research, and *politeness theory* research.

Research on virtual teams. Martins, Gilson & Maynard [82] recently surveyed the growing body of research on virtual teams (VT), which they defined as "teams whose members use technology to varying degrees in working across locational, temporal, and relational boundaries to accomplish an interdependent task" (p. 808). They found that the "majority of VT research pertaining to interpersonal processes… focused on conflict, uninhibited behavior…, informality of communication among group members, interpersonal trust, and group cohesiveness" (p. 814). Trust (one of the outcomes of group maintenance behavior) in particular has a rich literature. For example, Jarvenpaa and Leidner [61] identified what they called "swift trust" that formed in temporary distributed groups. However, Martins et al. note that much of this work has been done in a lab setting with student groups [82, p. 822], which is consistent with a focus on temporary teams. Such research needs to be followed up with studies of longer-standing functioning distributed groups, in particular because experience working together may be a key factor in developing relationships. They further note that "interpersonal processes represent an area in which major gaps exist in the literature on VTs." (p. 821), suggesting a need to consider the specific behaviors visible in CMC-mediated interaction that help build relationships.

Research in computer-mediated communications. To help identify these behaviors, we turn now to work that has examined CMC interaction in more detail. The notion of a *community of inquiry* has its antecedents in the work of the American pragmatists in general, and especially John Dewey [38, 89]. The term itself began to achieve wide usage through the work of Matthew Lipman and his Philosophy for Children movement [81]. A community of inquiry is characterized by trust and an open, critical, collaborative search for meaning and truth. No member is considered superior by virtue of organizational role or status. All members are expected to hear one another's ideas carefully, respond to them, correct them if necessary, and develop their own ideas without fear of harsh negative criticism or humiliation from others in the community [105]. In recent years, Anderson, Archer, Garrison and Rourke [3, 42, 43, 100] have developed and validated a content analysis scheme designed to evaluate the learning process of individuals using asynchronous technology to collaborate in a community of inquiry. Building on social interdependence, critical thinking, and constructivist learning theories [42, 51, 56, 86, 90, 115] they presented a model that integrates cognitive presence, social presence, and teaching presence. Their framework identifies the intellectual content of messages (*cognitive presence*), the instructional role (*teaching presence*), as well as the interaction among the members (*social presence*.) Aviv [6] also developed a framework to analyze the content of messages and the nature of interactions. His framework identifies three processes present in asynchronous learning network discussions: *social process*, *response process* and *reasoning process*. These frameworks provide a useful starting point for the identification of group maintenance behavior in asynchronous communication.

Another stream of research that provides useful insights into group maintenance behavior embedded in speech is *politeness theory*. Politeness theory considers the role of *face*, the positive self-image claimed and presented to the social world by each individual [45]. The theory posits that face-threatening acts (FTA) are an inherent and unavoidable aspect of any human interaction using language. Politeness in language represents an effort to support and preserve the self-esteem, or face, of others, to minimize the impact of face-threatening acts. Politeness tactics can be either specifically positive or negative [13]. Negative tactics attempt to avoid negative face by demonstrating distance and circumspection to the other [84]. Positive tactics indicate an appreciation of the other's wants in general [84]. Positive politeness tactics help group members to bond and to locate common ground whereas negative politeness tactics prevent group members from coming too close or intruding by keeping appropriate distance. Based on the work of Brown and Levinson [13], Morand and Ocker [84] developed a set of indicators of positive and negative politeness tactics for use in analyzing ICT transcripts.

We plan to build on both the community of inquiry and politeness theory frameworks in the proposed research because they identify markers in asynchronous communication that enhance the social dimension of computer-mediated collaboration. In other words, these indicators represent group maintenance behavior expressed in language, in contrast to earlier work that assumed face-to-face interactions. As an example, we present in Table 1 a preliminary set of group maintenance indicators identified for research on community of inquiry and politeness theory that we expect to see expressed in cyber-infrastructure supported communications. The table includes a range of indicators that make different tradeoffs between reliability and validity, i.e., some are very explicit and easy to recognize but perhaps only indirect indications of group maintenance, and vice versa. The indicators in the table represent a starting point for our research.

Antecedent and processes of group maintenance behavior. By developing a reliable and valid framework for identifying group maintenance behavior embedded in CMC, we can assess the role of specific forms of group maintenance behavior in the practices of the groups, looking for antecedents and processes of group maintenance. For example, one key question is which team members perform group maintenance behavior. In conventional groups, group maintenance behavior is often the role of team leaders, but in some of the groups we plan to study, formal indications of leadership are largely absent, leaving open the question of who, if anyone, takes responsibility for their performance.

Table 1. Initial constructs and indicators of group maintenance behavior**.**

Category	Indicators	Definition
Emotional expression [42, 100]	Expression of emotions	Conventional expressions of emotion, or unconventional expressions of emotion, includes, repetitious punctuation, conspicuous capitalization, emoticons.
	Use of humor	Teasing, cajoling, irony, understatements, sarcasm.
	Self-disclosure	Presents details of life outside of group activity, or expresses vulnerability
Interaction/open communication [42, 100]	Continuing a thread	Using reply feature of software, rather than starting a new thread.
	Quoting from others' messages.	Using software features to quote others entire message or cut and pasting selections of others' messages.
	Referring explicitly to others' messages [100]	Direct references to contents of others' posts.
	Asking questions	Ask others a question
	Complimenting; expressing appreciation [100]	Complimenting others or contents of others' messages.
	Expressing agreement	Expressing agreement with others or content of others' messages.
	Draw in participants [3]	Calling on other members to participate and including everyone in the discussion.
	Address individual member	Part of the message addresses a specific member(s)
Group cohesion	Phonological slurring [84]	Employ phonological slurring to convey in-group membership
	Colloquialism or slang [84]	Use colloquialism or slang to convey in-group membership
	Use ellipsis (omission) [84]	Use ellipsis (omission) to communicate tacit understandings
	Vocatives [100]	Addressing or referring to participants by name to insinuate familiarity
	Making personal connection [4]	Revealing commonalities with others in the group; raise or presuppose common grounds; express agreement
	Give reasons [84]	Assert reflexivity by making activity seem reasonable to the hearer.
	Inclusive	Use inclusive forms (we or lets) to include both speaker and hearer in the activity (we, our, us)
	Reciprocal exchange [84]	Assert reciprocal exchange or tit for tat
	Express sympathy, understanding	Give something desired: sympathy, understanding
	Apologies	Make apologies for doing something wrong
	Phatics, salutations [100]	Communication that serves a purely social function; greetings, closures.
	Encouraging others [4]	Encouraging others to do a work
	Common and symbolic language [3]	Members use of shared language/terms, analogies, symbols or metaphors specific to the group
Face protection [84]	Conventionally indirect	Be conventionally indirect; inquire into the hearer's ability or willingness to comply.
	Use hedges	Use hedges: words or phrases that diminish the force of a speech act.

	Subjunctive	Use subjunctive to express pessimism about hearer's ability/willingness to comply.
	Use words or phrases that minimize the imposition.	Use words or phrases that minimize the imposition.
	Honorifics	Give deference by using honorifics: Sir, Mr., Ms., Dr.
	Formal word choices	Use formal word choices to indicate seriousness and to establish social distance.
	Apologies	Apologize: admit the impingement, express reluctance.
	Impersonalization	Impersonalise the speaker and hearer by avoiding the pronouns "I" and "you."
	Past tense	Use the past tense to create distance in time.
	Nominalization	Nominalize (change verbs & adverbs into adjectives or nouns) to diminish speakers' active participation.
	General rule	State a face-threatening act as a general rule.

Group outcomes. Finally, we plan to evaluate the relation of group maintenance behavior to group effectiveness. Research has empirically linked group maintenance behavior in face-to-face groups with several indicators of positive group or organizational performance. For example, organizational citizenship behavior has been associated positively with performance quantity, performance quality, financial efficiency, customer service, and attitudes such as job satisfaction [94]. There are a number of reasons why group maintenance behavior may contribute positively to group performance. Hackman [50] and Gladstein [44] argue that these behaviors satisfy the social needs of group members and contribute to group synergy. Aviv [6] and Rourke et al., [100] suggested that cohesiveness and positive interpersonal characteristics in a group promote information sharing and learning. Such behavior may also improve the group's ability to attract and retain high-quality members, and may improve performance by enhancing morale, group cohesiveness, job satisfaction, and the sense of belonging to a group [94]. Because much of the research linking group maintenance behavior to positive outcomes has been cross-sectional and correlational in nature, it is difficult to make conclusive assertions about causality. Nevertheless, research continues to accumulate in support of a positive association.

For this research, we will consider effectiveness along the three dimensions suggested by Hackman [50] as shown in Figure 1 above: task performance, as measured by evaluations by recipients of the output, individual group member satisfaction and continued group performance. For the FLOSS setting in particular, Crowston et al. [26] have developed a set of indicators of effectiveness, including releases and bug fixes as measures of task performance, individual developer satisfaction with the project, and number of developers involved and level of activity as indicators of continued group performance. We anticipate that the effects of group maintenance behavior will be more visible in certain of these outcomes, e.g., we expect it to have a large impact on the group's ability to retain members, though the nature of and mechanisms for the relationship are the subject of the proposed study. We plan to adapt similar measures to evaluate other kinds of distributed groups.

2. Research Design

In this section, we discuss the design of the proposed study, addressing the basic research strategy, concepts to be examined, sample populations and proposed data collection and analysis techniques. We first discuss the goals and general design of the study. We then present the details of how data will be elicited and analyzed.

We envision our project as having two overlapping phases for each of the group domains studied. Each phase will last roughly a year, though the transition between these phases will be gradual rather than a sharp boundary. In the first phase (roughly year 1), we will use computer-assisted qualitative data analysis software (CAQDAS) to examine project data (e.g., e-mail archives, computer logs, primary and secondary documents and source material) manually for evidence of the concepts identified in Table 1 to determine what kinds of group maintenance seem most important in our groups and to elucidate the connection between group maintenance behavior and antecedents and outcomes in order to propose more specific hypotheses for further study. We will also examine how these concepts are linguistically realized in text in order to determine feasible candidates for identification using NLP techniques, by delineating the predictable linguistic features on which algorithms to detect the research-relevant features can be based. In the second phase (roughly year 2), we will use the NLP algorithms to extract larger numbers of the identified research-relevant features from the datasets and use these techniques to carry the analysis across a larger number of projects. This approach will allow us to more fully explore the large-scale digital datasets developed in the project, to compare the NLP results with the CAQDAS results and to gauge the generalizability of the proposed hypotheses.

2.1 Sample

We will start each phase by identifying promising distributed groups for study. During the first phase, we will focus on a small number of groups (on the order of six). In the second phase, the size of the sample will be limited by the available data and processing power (computer and human). In choosing these groups we will apply the previously developed effectiveness assessments (described above) as a theoretical sampling filter to ensure that we have groups of different types with varying degrees of effectiveness. We will also take into consideration some pragmatic considerations, such as selecting only projects where we have access to the needed data. We plan to identify distributed groups in three domains: 1) scientific research collaboratories; 2) transnational policy networks; and 3) FLOSS software development groups. These types of groups have been chosen because each involves collaborations between geographically and organizational separated members, carried out primarily via cyber-infrastructure, in order to accomplish shared tasks that produce some kind of innovation. In the remainder of this section, we discuss each form in turn.

Scientific Research Collaboratories. The first form of distributed group we plan to study is a scientific research collaboratory. The collaboratory concept combines the words collaborate and laboratory to refer to “a center without walls” in which the nation’s researchers could be geographically distributed and yet collaborate as if they were in the same physical location [117]. More recently, the definition of research collaboratories has been refined to see it as “an organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and provides access to data sources, artifacts and tools required to accomplish research tasks” [39]. The number of collaboratories has increased immensely since their inception in the 1980’s [40, 41], as the concept has been applied in settings from education to astrophysics, from genomics to manufacturing [15-17, 92]. Much of the research on collaboratories has focused on designing and developing the technologies or the technical infrastructures for the collaboratories [e.g., 8, 14, 107], so a study of the social aspects will be a contribution to this area. Because of our interest in cyber-infrastructure-supported distributed groups, we plan to focus our attention on research collaboratories that rely primarily on ICT to support interactions (i.e., as opposed to primarily working via periodic face-to-face meetings). The NSF funded Science of Collaboratories (SOC, http://www.scienceofcollaboratories.org/) project has compiled an inventory of collaboratories of various types, and provides summaries, links to their websites and detailed analysis of a limited number of collaboratories. Using this inventory as a sampling frame, for Phase I we plan to draw a small purposive sample of 2 collaboratory projects that are willing to provide data for the study, or who make their data publicly available. Ideally, we will identify collaboratories working in similar areas to ensure comparability. For example, NIH has funded a number of collaboratories in structural genomics. We have attached letters of support from leaders of two such collaboratory, the Northeast Structural Genomics (NESG) Consortium and the Berkeley Structural Genomics Centre, indicating their interest in our research question and willingness to negotiate a working relationship. In Phase II, we will expand the study to include as many collaboratories as reasonable, given the available time and resources of the project.

Transnational Advocacy Networks. The second kind of distributed group we plan to study are Transnational Advocacy Networks (TANs), such as those associated with the recently concluded United Nations World Summit on the Information Society (WSIS). TANs are distributed groups of individuals and organizations with the shared goal of influencing national or international policy. In the case of WSIS, its structures were designed to explicitly involve governments, the private sector, and civil society. By definition, the WSIS civil society is transnational in scope, with hundreds of individuals (sometimes representing organizations) participating in one or more of its many organically emerging structures [15]. At last count, the WSIS civil society had the following major self-organizing components: the civil society bureau; the civil society plenary; the content and themes group; and a further twenty-two working groups, caucuses and task forces. Each of these components involves the collaboration of a geographically distributed group. While some prominent members of these TANs do meet face-to-face periodically during preparatory meetings for the WSIS or related conferences, their work is supported primarily by CMC tools, specifically e-mail lists [15]. For this aspect of the project, we plan in Phase I to focus on two of these groups. The first is the WSIS Civil Society Plenary (CSP), which is seen within the civil society structures as the most "legitimate" structure; however, it is also the largest and most unwieldy of the various structures. The second is the Internet Governance Caucus (IGC). This caucus is important because of its significant input into the most important policy debate of the WSIS processes, which centered around a transformation of the international regime to provide global governance for the Internet. Again, in Phase II, we will expand the study to include other TANs.

Free/Libre Open Source Software (FLOSS) development groups. Our final set of distributed groups are FLOSS development teams. FLOSS is a broad term used to embrace software developed and released under an "open source" license allowing inspection, modification and redistribution of the software's source[1]. There are thousands of FLOSS projects, spanning a wide range of applications. Due to their size, success and influence, the Linux operating system and the Apache Web Server and related projects are the most well known, but hundreds of others are in widespread use, including projects on Internet infrastructure (e.g., sendmail, bind), user applications (e.g., Mozilla, OpenOffice) and programming languages (e.g., Perl, Python, gcc) and even enterprise systems (e.g., eGroupware, Compiere, openCRX). Key to our interest is the fact that most FLOSS software is developed by self-organizing distributed groups comprising professionals, users [109-111] and other volunteers working in loosely-coupled groups. These groups are close to pure virtual groups in that developers contribute from around the world, meet face-to-face infrequently if at all, and coordinate their activity primarily using a cyber-infrastructure [97, 113]. The groups have a high isolation index [87] in that most group members work on their own and in most cases for different organizations (or no organization at all). While these features place FLOSS groups at one end of the continuum of distributed work arrangements, the emphasis on distributed work makes them useful as a research setting for isolating the implications of this organizational innovation. For Phase I, we will chose two projects that produce comparable systems in order to control for the nature of the program, thus allowing a more direct comparison of the groups' effectiveness. For example, in other work, we have compared two Internet Messaging clients. Again, in Phase II, we will expand the study to include a larger diversity of projects.

2.2 Data collection and cleaning

To explore the concepts identified in the conceptual development section of this proposal (Table 1), we will collect and analyze a range of data (e.g., e-mail archives, computer logs, primary and secondary project source documents and possibly supplemented with interviews with members of the initial projects). The most voluminous source of data will be collected from archives of ICT tools used to support the groups' interactions [54, 66]. These data are useful because they are unobtrusive measures of the group's behaviors [114]. In particular, mailing list archives will be a primary source of interaction data that illuminates the role of social maintenance, as email is one of the primary tools used to support group communication [65]. In the FLOSS setting, such archives are the primary mode of communication and so contain a huge amount of data (e.g., the Linux

---

1 FLOSS software is usually available without charge ("free as in beer"). Much (though not all) of this software is also "free software", meaning that derivative works must be made available under the same unrestrictive license terms ("free as in speech", thus "libre"). We have chosen to use the acronym FLOSS rather than the more common OSS to acknowledge this dual meaning.

kernel list receives 5-7000 messages per month, the Apache httpd list receives an average of 40 messages a day). The TANs involved with the WSIS civil society also have large archives of email available for analysis (e.g., more than 8000 messages over 18 months for the WSIS Civil Society Plenary). Most scientific collaboratories also rely on ICT for day-to-day interactions, so we anticipate being able to identify comparable sources of data.

We will download this data from the message archives, clean the data (e.g., by removing unnecessary coding from attachments), provide descriptive metadata on each archive, and extract the date, sender and any individual recipient' names, the sender of the original message, in the case of a response, and text of each message. We will also code additional demographic data (where available) such as gender, region, organization and role within the group. Some of this manual coding will be facilitated by limited auto-coding within the CAQDAS tools used in the study (e.g., Atlas/ti or Hyper-research).

2.3 Data analysis

While voluminous, the data described above are at a low level of abstraction. The collected data will be analyzed using a variety of techniques to raise the level of conceptualization to fit our theoretical perspective.

Phase I. In phase I, we will use CAQDAS tools to conduct content analysis to reduce the large amount of raw data to more specific codes and measures. Content analysis of computer-mediated communication has been an active area of research [10, 55]. Data will be content analyzed following the process suggested by Miles and Huberman [83], iterating between data collection, data reduction (coding), data display, and drawing and verifying conclusions—increasingly supported by the evolving NLP content analysis tools. The initial (deductive) framework will be based on the conceptual development reviewed above, but we plan to evolve this framework based on our experiences with the data. As such, the research will also engage open code instances of group maintenance and other social support mechanisms. In addition, this phase will allow us to develop hypotheses about the relationship between group maintenance behavior and group performance across various settings, based on a developing understanding of the processes of group maintenance and its role in the life of the groups. A proportion of messages (ideally 100%) will be coded by two individuals to enable calculations of reliability.

As an example, the content analysis approach was applied to one FLOSS project, Apache httpd, in a PhD thesis by a student in our group at Syracuse [4]. Though her main focus was on group learning, Annabi identified some instances of group maintenance behavior that affected the social atmosphere in the group. For example, developers almost always referred to the group with inclusive pronouns and addressed each other by name. They shared limited personal stories (e.g., wedding and honeymoon news) such as:

```
"If someone could implement this since I *swear* this is
the last time I'm logging in before the wedding (haha) feel
free—at the very least comment out the code relating to
"experts" until that's implemented."
```

Humor and other forms of emotional expression were also common (e.g., "oop ack!" or "RTF owes RH beer"). There was little 'flaming'—escalating email hostility—in group interactions, even when frustrations and strong feelings were shared. In Annabi's study, group maintenance behavior were performed by all members on a small scale, instead of just by the group leader. Only in conflict resolution did the two leaders of the group show higher levels of group maintenance behavior. In another example, a similar approach to content analysis, albeit using different constructs, was taken by a doctoral candidate in the School of Information Studies studying a sub-set of the WSIS plenary e-mail archive for her doctoral dissertation. Although not yet completed, Zakaria has been able to develop some important findings on the impact of communication style on decision-making in the WSIS civil society [122]. She has also been able to develop innovative methods for training the coders, to test for inter-coder/rater reliability and to manage remote access to the coding database and software, techniques that will be employed in the management of this larger project. These examples demonstrate two things: first, the feasibility of our planned approach and second, the nature of the phenomenon we will identify. However, studies of a single project cannot identify how these phenomenon relate to project effectiveness, hence the need for the proposed study.

Phase II. In the second phase of the project, we will utilize and extend current NLP technology to assist in identifying important semantic patterns that can then be translated into the codes developed in Phase I. Turner et al. [108] similarly used some simple NLP approaches to analyze bug reports, though our proposed work goes well beyond this initial effort. Use of NLP techniques supports researchers in looking for patterns with greater reliability and across larger amounts of data and a wider range of projects. These techniques can identify patterns that would be too tedious for human coders (e.g., frequency of particular words), or more subtle indications of implicitly conveyed concepts (e.g., the initiator of particular kinds of behavior). In some cases, the NLP techniques can replace manual coding, but are also useful in combination with human coding. For example, NLP techniques might be used to quickly identify candidate episodes of the use of humour (one of our indicators), which a human coder could then qualitatively assess. Because the use of NLP techniques is one of the major innovations of this proposal and the foundation for further analysis, we will explain its application in more detail.

Application of NLP-based text processing to CMC transcripts (e.g., chat room conversations or emails) has been a challenge given the nature of these interactions. These texts are known for their use of specialized language patterns, as well as informal grammar and spelling rules [96]. To effectively meet the challenge of understanding these stylistically diverse and grammatically inconsistent texts, our NLP technology will leverage theoretical and empirical advances in research on Sublanguage Analysis and Discourse Structure. A sublanguage is defined as the particular language usage patterns that develop within the written or spoken communications of a community that uses this sublanguage to accomplish some common goal or to discuss topics of common interest. The fact that a sublanguage deals with a restricted domain and is used for a specific purpose results in useful restrictions on the range of linguistic data that needs to be accounted for by the system. At the lexical level, the sublanguage excludes large parts of the total vocabulary of a language; for those words in the sublanguage vocabulary, the number of senses actually used for each word is limited. At the syntactic level, a sublanguage is characterized by predictable surface structures, utilizes a limited range of verbs, and makes extensive use of domain-specific nominal compounds, which reflect the specialized nature of the sub-field. The discourse level of a sub-language deals holistically with units of language larger than a sentence, relying on the predictable structure of communications in this sublanguage. The discourse level model of a particular communication type consists of semantic categories (reflecting the purpose of communication) and the relations among those categories. The NLP system's recognition of these semantic categories handles the great surface variety in terms of lexical and syntactic choices in how entities (e.g., people, organizations), events (e.g., updates, requests), and relations amongst them (e.g., who requests an action by whom) are realized in text. As a result, the sublanguage analysis is able to abstract up from these individual instances that indicate presence of the underlying concepts to reveal the features of the model (e.g., politeness, community of inquiry) under study. Communication types that have been analyzed and for which sublanguage grammars have been developed include abstracts, news articles, arguments, instructions, manuals, dialogue, instructions, email, and queries [80]. Early research in Sublanguage Theory [49, 79, 80, 101] has shown that there are recognizable linguistic differences amongst various types of discourse (e.g., news reports, email, manuals, requests, arguments, interviews) and that discourses of a particular type that are used for a common purpose within a group of individuals exhibit characteristic linguistic (lexical, syntactic, semantic, discourse, and pragmatic) features. Humans use these characteristic features to extract meaning, and these human processes can be simulated by a full-fledged NLP system in order to extract levels of meaning beyond the simple surface facts.

In the proposed research, the sublanguage analysis framework will be applied to automatically identify the important linguistic patterns in the text-based electronic communications that will be processed by the NLP system, and to annotate them with initial content categories, which will then be refined by the project group to reflect the conceptual framework emerging from the data. The

NLP-based software developed at the Center for Natural Language Processing (CNLP) at Syracuse University analyzes naturally occurring texts (e.g., documents, transcribed interviews, email, chat) for the explicit and implicit meanings which are conveyed (and which usually only a human would recognize). The resulting NLP annotations will be used as initial codes representing the items such as the events, roles, intentions, goals or expectations reported and/or hinted at in the text (e.g., names, popular abbreviations, special terms, time expressions and other phrases with particular semantic values relevant to the research agenda). Some of the group maintenance features identified in section 2 are explicit and thus straightforward to identify (e.g., use of inclusive pronouns or past tense); others are more implicit (e.g., use of humour, self-disclosure or emotion) and will thus require careful analysis and extension of the current NLP tools to identify. (It should be noted that such features are often also problematic for human coders to reliably identify.)

As an example of this approach, we have conducted a small pilot study using CNLP's tools in a semester-long doctoral seminar, wherein students wrote rules to identify the politeness behaviors (or strategies) used in two FLOSS projects: Bibdesk, a graphical bibliography manager for the Mac OS X operating system, and GAIM, a multi-platform instant messaging and Inter-Relay Chat application. The students analyzed nearly 10,000 email messages from the Bibdesk and GAIM developers' list for evidence of both positive and negative politeness behavior (i.e., using just one of the frameworks discussed above). They made progress on identifying several of the indicators, such as politeness and hedges, but found that indicators such as personal connection or impersonalization were much more difficult. This initial effort demonstrates that the NLP techniques do have promise for coding group maintenance behavior expressed via cyber-infrastructure. However, this initial effort has clearly only scratched the surface of what is possible, and demonstrates the need for more intensive effort in specializing the NLP technology to capture more complex group maintenance behavior, as well as the need to test the rules on a variety of different datasets.

The tools developed by CNLP (e.g., Vanilla Extract, Knowledge Base Builder) for use by human coders will support a positive synergy between the manual and NLP coding, as each suggests indicators for further analysis. However, we also anticipate learning from the differences in coding approaches. For example, with human coders, coding reliability must be assessed by double coding, while automated coding offers the possibility of 100% reliability. On the other hand, we will need to carefully examine what is and is not coded to assess the validity of the automated coding (though this is an issue with manual coding as well). Another example of a difference in approach is in the choice of unit of coding. In manual coding, it is common to use the semantic unit as the unit of coding, while for automated coding, the unit of coding needs to be unambiguously identifiable, e.g., the sentence or message.

3. Management plan

Based on preliminary assessment of the effort required, we are requesting funding for two graduate students and a small amount of summer support for 4 PIs (approximately 0.4 summer months per PI). All four PIs, Drs. Derrick L Cogburn, Kevin Crowston, Robert Heckman and Elizabeth D. Liddy, will work during the summer on project management and research design, and devote 10% of effort during the academic year to project management and oversight (1/2 day per week, supported by Syracuse University). All four PIs will share in project selection, overall project design and report writing. Each PI will be responsible for designing specific aspects of the project and overseeing work on those aspects:

- Dr. Crowston will direct the project and be responsible for general project oversight and reporting to NSF.
- Drs. Heckman and Crowston will lead the research on the FLOSS groups.
- Dr. Cogburn will lead the research on scientific collaboratories and transnational policy networks.
- Dr. Liddy will lead the computer/information science research team in NLP tool development and integration. Dr. Liddy has extensive leadership experience in successful delivery of innovations in NLP based on 60+ funded R&D projects.

The graduate students will devote 50% effort during the academic year and 100% effort during the summers, for a total of 2200 hours/year (4400 hours in two years). The graduate students will support the principal investigators in sample section, definition of constructs and variables, and will have primary responsibility for data collection and analysis, under the oversight of the PIs. Manual content analysis is extremely labor intensive. Our experience using this technique in other projects suggests that a student working with a PI will be able to code at most two projects worth of messages in a semester. For Phase I, two students working for one full year should be able to double code the six projects we plan to analyze in the Phase I (though if necessary, we can double code just a subset of messages to assess the reliability of the code book and single code the remaining messages).

For Phase II, we anticipate having one student devote full time working with a PI on developing NLP rules for identifying group maintenance behavior while the second student supports running projects through the system and manually coding projects in conjunction with the automated coding. Note that Phase I and Phase II will partially overlap, meaning that we will finish the manual coding of the six initial groups in parallel with developing NLP rules. We will start Phase I by working on the TAN and FLOSS groups, for which we already have data, and continue with the research collaboratories as the data becomes available. A time line is included as part of the budget justification to show how the requested resources will be employed. We are also applying for support from NSF for an IGERT grant, which if successful will provide resources for additional doctoral students to get involved in this innovative interdisciplinary research project and to learn the theory and practice of distributed collaboration.

We will employ two main project management techniques. First, we will have regular meetings of the project members to share findings and to plan the work. Initially, these will be every other week, but the frequency of meetings will be adjusted depending on our experience and the pace of the work being carried out at the time. These formal meetings of all project participants will augment the regular interaction of the teams of PIs and students working on the data analysis and expected frequent interactions of the students as they analyze data from the same projects. The NLP development team, all of whom are co-located in CNLP, will meet semi-weekly during the design phases and then weekly during implementation. The experience of this team on the existing toolset bodes well for an accelerated process of iterative requirements, implementation, usage, and new requirements. Second, an initial project activity will be the development of a more detailed timeline (based on the initial one in the budget justification section) against which progress will be measured. The budget includes support for PIs and PhD students during summer and academic year to support these activities.

4. Conclusion

In this proposal, we develop a conceptual framework and a research plan to investigate group maintenance functions within distributed groups, using a combination of manual and NLP content analysis of interaction carried out via cyber-infrastructure. The proposed project will have both intellectual and broader impacts.

Expected intellectual merits

The project will contribute to advancing knowledge and understanding of distributed groups by identifying the role of group maintenance for distributed groups. We expect this study to make conceptual, methodological as well as practical contributions. Understanding the role of group maintenance in a group of independent knowledge workers working in a distributed environment is important to improve the effectiveness of distributed groups and of the traditional and non-traditional organizations within which they exist. Developing a theoretical framework consolidating a number of theories to understand the role of group maintenance behavior within a distributed group is an important contribution to the study of distributed groups.

Expected broader impacts

The project has numerous broader impacts. The project will benefit society by identifying the role of group maintenance in distributed groups, focusing on groups that are responsible for developing innovative outcomes. Such groups are an increasingly important approach to needs such as software development, scientific research and policy development. Understanding the role of group maintenance in these settings and the relation to group performance will help us develop guidelines to improve performance and foster innovation. Distributed work groups potentially provide several benefits but the separation between members of distributed groups creates difficulties in building social relations, which may ultimately result in a failure of the group to be effective. For the potential of distributed groups to be fully realized, research is needed on how to make these groups engaging and motivating to members.

To ensure that our study has a significant impact, we plan to broadly disseminate results through journal publications, conferences, workshops and on our Web pages, as well as through our interaction with the leaders and members of distributed teams. Our results could also be incorporated into the curricula of the professional degrees of the Syracuse University School of Information Studies, as well as improving the pedagogy of our courses and degree programs, as these programs are offered on-line and thus involve distributed groups. The project will promote teaching, training, and learning by students in the research project, providing them the opportunity to develop skills in model development, theory application, data collection and analysis.

Results from prior NSF funding

Three of the PIs for this grant, Drs. Crowston, Heckman and Liddy, have been jointly funded by one NSF grant within the past 48 months, HSD 05–27457 ($684,882, 2005–2008), *Investigating the Dynamics of Free/Libre Open Source Software Development Teams*. This project was funded at the end of 2005, and work on it has just begun. The current proposal differs from the HSD project in its focus on the role of social maintenance rather than task-oriented behavior and its inclusion of other kinds of distributed groups, in particular, scientific and policy collaboratories supported by cyber-infrastructure. Nevertheless, the PIs' experience working together will be beneficial for the management of the current proposal. As well, we expect substantial synergies between the projects that will facilitate the proposed research.

Dr. Crowston has been funded by an additional four NSF grants in the past 48 months. The two most closely related to the current proposal are IIS 04–14468 ($327,026, 2004–2006) and SGER IIS 03–41475 ($12,052, 2003–2004), both entitled *Effective work practices for Open Source Software development*. These grants have provided support for travel to conferences (e.g., *ApacheCon* and *OSCon*) to observe, interview and seek support from developers and to present preliminary results, and for the purchase of data analysis software and equipment. This work has resulted in five journal papers [24-26, 32, 58], multiple conference papers [e.g., 19, 22, 27, 34, 35, 59] and workshop presentations [e.g., 18, 20, 21, 23, 33, 60], with additional papers under review. These grants support a total of four PhD students; several others have been involved in aspects of the work.

Crowston's fourth grant is IIS 04–14482 ($302,685, 2005–2006, with Barbara Kwasnik), for *How can document-genre metadata improve information-access for large digital collections?* The grant partially supported work on conference papers, a conference mini-track and journal special issue [64]. Earlier work by the PIs on genre has appeared in journals [e.g., 28] and conference papers [e.g., 63]. The grant funds two PhD students; two others are involved in aspects of the research. Earlier support came from IIS–0000178 ($269,967, 2000–2003), entitled *Towards Friction-Free Work: A Multi-Method Study of the Use of Information Technology in the Real Estate Industry*. The goal of that study was to examine how the pervasive use of information and communication technologies (ICT) in the real-estate industry changes the way people and organizations in that industry work. The project resulted in several journal articles [29, 31, 37, 85, 102, 103] and numerous conference presentations [e.g., 30, 36]. One PhD student is finishing a thesis based on this work.

The Co-PI of this proposal, Dr. Liddy, has received NSF funding for six projects in the past five years. One is listed above (with Crowston and Heckman), a second is briefly described here, while the remaining four form a cohesive research program, which is described in more detail below. The second grant was DUE-0241856 ($2,519, 166, 2002-06), entitled *Multidisciplinary Systems Assurance Education.* As a Co-PI on this Federal Cyber Service Scholarship for Service Program grant, Liddy serves as mentor for Masters students in both the Information Management and Telecommunication & Network Management degree programs. Liddy has been able to involve the students actively in appropriate funded research projects underway at CNLP that address issues of information systems security and insider threats. Efforts have been established to enable these students to have summer internships with local companies whose expertise is in R & D on information systems security. The contribution to human resources is the main goal of this project and it is showing promising results.

The remaining four projects are funded by NSF's National Science Digital Library Program and involve research, implementation, and evaluation of NLP technology for automatic metadata generation for educational objects, most typically teachers' lesson plans and activity guides. The four grants are DUE-0085837 ($366,000, 2000-02) *Breaking the MetaData Generation Bottleneck*, DUE-0121543 ($475,000, 2001-03), *Standard Connection: Mapping Educational Objects to Content Standards*, DUE-0226312 ($374,938, 2002-04), *MetaTest: Evaluating the Quality & Utility of MetaData* and DUE-0435339 ($634,218, 2004-2006*), Computer-Assisted Standard Assignment & Alignment*. Two of the projects center around content standards, either their automatic assignment to resources or the automatic mapping amongst multiple national standards and the fifty state standards. Over the life of these four projects, Liddy and group have: 1) adapted their existing NLP methods and technology to the task of extracting from learning resources the values for the 23 metadata elements used for representing learning objects in digital libraries (15 Dublin Core + 8 GEM); 2) proven in end-user empirical evaluations that the metadata elements assigned automatically using NLP are equally good as those assigned by humans, and; 3) extended the metadata capability to map individual resources to the relevant content standards in Math and Science, key to standards-based education, and automated state-to-state and state-to-national alignment of content standards. Results were evaluated by experts in standards and by classroom teachers. The grants have resulted in numerous publications [67-78, 119]. Four PhD students and three Masters students have been active participants, learning both about the research and evaluation process and the wider field of digital libraries. They have presented the projects' findings jointly or singly and interacted substantively with this research community at relevant conferences.

The fourth co-PI on this proposal, Dr. Cogburn, has received NSF funding for 3 projects in the past 5 years. The most recent project, *Transnational Non-Governmental Organizations and Dynamic Change*, NSF HSD 05–27679, ($500,000, 2005-07, with Margaret Herman), was awarded in the fall of 2005. This grant supports a large-scale, systematic study of transnational Non-Governmental Organizations (TaNGOs) as agents of social change. Using a concurrent mixed-methods design (e.g., interviews, surveys, comparative case studies, and web-based archival research), the study is investigating the impact of leadership, structure, communication, and collaboration on the effectiveness, and accountability of these organizations at national, regional, and international levels. This project will provide synergies for the proposed study of TANs.

Dr. Cogburn served as senior personnel on two NSF grants while at the University of Michigan. The first, ITR/SOC+IM-0085951 ($2,400,000, 2001-05), *Sustainable and Generalizable Technologies to Support Collaboration in Science*, focused on identifying socio-technical factors that affect the development of scientific collaboratories, and ways to apply those lessons to new collaboratory development (see http://www.scienceofcollaboratories.org/). Ten publications and numerous scholarly presentations were supported in part by his association with the SOC project. Knowledge from these projects informs the research in this proposal. The second grant, IGERT-0114368 ($2.7 Million, 2001-05) entitled *Socio-Technical Infrastructure for Electronic Transactions (STIET)* (http://www.si.umich.edu/stiet/index.htm), explored the institutional and technical mechanisms that support the development of electronic commerce. For this project, Dr. Cogburn mentored doctoral students participating in the STIET program and participated in weekly seminars. Knowledge from the project strengthened the approach to the socio-technical infrastructure to support geographically distributed collaboration now used within Cogburn's research group.

References

[1] M. K. Ahuja, D. F. Galletta, and K. Carley, “Individual performance in virtual groups,” *Management Science*, vol. 49, pp. 21–38, 2003.

[2] D. G. Ancona and D. F. Caldwell, “Beyond task and maintenance: Defining external functions in groups.,” *Group and Organization Studies*, vol. 13, pp. 468–494, 1988.

[3] T. Anderson, L. Rourke, W. Archer, and R. Garrison, “Assessing teaching presence in computer conferencing transcripts,” *Journal of the Asynchronous Learning Network*, vol. 5, 2001.

[4] H. Annabi, “Moving from Individual Contribution to Group Learning: The Early Years Of The Apache Web Server,” in *School of Information Studies*. Syracuse, NY: Syracuse University, 2005.

[5] D. E. Atkins, Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P., & Wright, M.H., “Revolutionizing science and engineering through cyberinfrastructure: Report of the Blue-Ribbon Advisory Panel on Cyberinfrastructure,” National Science Foundation, Washington, DC 2003.

[6] R. Aviv, “Educational performance of ALN via content analysis,” *The Journal Of Asynchronous Learning Networks*, vol. 4, 2000.

[7] R. Axelrod, *The evolution of co-operation*. New York: Basic Books, 1985.

[8] G. Bafoutsou and G. Mentzas, “Review and functional classification of collaborative systems,” *International Journal of Information Management*, vol. 22, pp. 281–305, 2002.

[9] R. F. Bales, “The equilibrium problem in small groups,” in *Working papers in the theory of action*, T. Parsons, R. F. Bales, and E. A. Shils, Eds. Glencoe, IL: Free Press, 1953, pp. 111–161.

[10] M. Beißwenger, “Bibliography of Chat Communications,” vol. 2004, 2003.

[11] E. Bender, “Rules of the Collaboratory Game,” in *Technology Review*, vol. 23 November, 2004.

[12] F. B. Berman, H, “Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. Available at www.sdsc.edu/sbe/,” National Science Foundation 2005.

[13] P. Brown and S. Levinson, *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press, 1987.

[14] B. R. Cleal, H. H. K. Andersen, and H. Albrechtsen, “Collaboration, communication and categorical complexity: A case study in collaboratory evaluation,” *Journal of Digital Information*, vol. 2, pp. 13–20, 2004.

[15] D. L. Cogburn, “Diversity matters, even at a distance: Evaluating the impact of computer-mediated communication on civil society participation in the World Summit on the Information Society,” *Information Technologies and International Development*, vol. 1, pp. 15–42, 2004.

[16] D. L. Cogburn, & Levinson, N., “US-Africa virtual collaboration in globalization studies: Success factors for complex, cross-national learning teams,” *International Studies Perspectives*, vol. 4, pp. 34-51, 2003.

[17] D. L. Cogburn and L. Zhang, “Still Going Global, Locally: The Impact of Team Mode in a Distributed Collaborative Learning Environment (Extended and updated version),” *IADIS International Journal on WWW/Internet*, vol. 1, 2004.

[18] M. Conklin, J. Howison, and K. Crowston, “Collaboration Using OSSmole: A repository of FLOSS data and analyses,” in *Symposium on Mining Software Repositories*. St. Louis, 2005.

[19] K. Crowston, H. Annabi, and J. Howison, "Defining Open Source Software project success," presented at Proceedings of the 24th International Conference on Information Systems (ICIS 2003), Seattle, WA, 2003.

[20] K. Crowston, H. Annabi, J. Howison, and C. Masango, "Effective work practices for Software Engineering: Free/Libre Open Source Software Development," presented at WISER Workshop on Interdisciplinary Software Engineering Research, SIGSOFT 2004/FSE-12 Conference,, Newport Beach, CA, 2004.

[21] K. Crowston, H. Annabi, J. Howison, and C. Masango, "Towards a portfolio of FLOSS project success measures," presented at Workshop on Open Source Software Engineering, 26th International Conference on Software Engineering, Edinburgh, 2004.

[22] K. Crowston, R. Heckman, H. Annabi, and C. Masango, "A structurational perspective on leadership in Free/Libre Open Source Software teams," presented at OSSCon, Genova, Italy, 2005.

[23] K. Crowston and J. Howison, "The social structure of Open Source Software development teams," presented at The IFIP 8.2 Working Group on Information Systems in Organizations Organizations and Society in Information Systems (OASIS) 2003 Workshop, Seattle, WA, 2003.

[24] K. Crowston and J. Howison, "The social structure of free and open source software development," *First Monday*, vol. 10, 2005.

[25] K. Crowston and J. Howison, "Hierarchy and Centralization in Free and Open Source Software team communications," *Knowledge, Technology & Policy*, In press.

[26] K. Crowston, J. Howison, and H. Annabi, "Information systems success in Free and Open Source Software development: Theory and measures," *Software Process—Improvement and Practice*, In press.

[27] K. Crowston, J. Howison, C. Masango, and U. Y. Eseryel, "Face-to-face interactions in self-organizing distributed teams," presented at Academy of Management Conference, Honolulu, HI, 2005.

[28] K. Crowston and B. H. Kwasnik, "Can document-genre metadata improve information access to large digital collections?," *Library Trends*, vol. 52, pp. 345–361, 2003.

[29] K. Crowston and M. D. Myers, "Information technology and the transformation of industries: Three research perspectives," *Journal of Strategic Information Systems*, vol. 13, pp. 5–28, 2004.

[30] K. Crowston, S. Sawyer, and R. Wigand, "Investigating the interplay between structure and technology in the real estate industry," presented at Organizational Communications and Information Systems Division, Academy of Management Conference, Chicago, IL, 1999.

[31] K. Crowston, S. Sawyer, and R. Wigand, "Investigating the interplay between structure and technology in the real estate industry," *Information, Technology and People*, vol. 14, pp. 163–183, 2001.

[32] K. Crowston and B. Scozzi, "Open source software projects as virtual organizations: Competency rallying for software development," *IEE Proceedings Software*, vol. 149, pp. 3–17, 2002.

[33] K. Crowston and B. Scozzi, "Coordination practices for bug fixing within FLOSS development teams " presented at Presentation at 1st International Workshop on Computer Supported Activity Coordination, 6th International Conference on Enterprise Information Systems, Porto, Portugal, 2004.

[34] K. Crowston, K. Wei, Q. Li, U. Y. Eseryel, and J. Howison, "Coordination of free/libre open source software development," presented at International Conference on Information Systems, Las Vegas, 2005.

[35] K. Crowston, K. Wei, Q. Li, and J. Howison, “Core and periphery in Free/Libre and Open Source software team communications,” presented at Hawai'i International Conference on System System (HICSS-39), Kaua'i, Hawai'i, 2006.

[36] K. Crowston and R. Wigand, “Use of the web for electronic commerce in real estate,” presented at Association for Information Systems Americas Conference, Baltimore, MD, 1998.

[37] K. Crowston and R. Wigand, “Real estate war in cyberspace: An emerging electronic market?,” *International Journal of Electronic Markets*, vol. 9, pp. 1–8, 1999.

[38] M. Daniel, “John Dewey and Matthew Lipman: Pragmatism in education,” in *Children: Thinking and Philosophy*, D. Camhy, Ed. Sankt Augustin: Academia Verlag, 1994.

[39] T. Finholt, G. Olson, J. Olson, and S. Teasley, “Workshops : Comparative Investigation : Final Summary,” vol. 2006. Ann Arbor, MI, 2002.

[40] T. A. Finholt, “Collaboratories,” in *Annual Review of Information Science and Technology*, vol. 36, B. Cronin, Ed. Medford, NJ: Information Today, 2001.

[41] T. A. Finholt and G. M. Olson, “From laboratories to collaboratories: A new organizational form for scientific collaboration,” *Psychological Science*, vol. 8, pp. 28-35, 1997.

[42] R. Garrison, T. Anderson, and W. Archer, “Critical thinking in a text-based environment: Computer conferencing in higher education,” *The Internet and Higher Education*, vol. 2, pp. 87-105, 2000.

[43] R. Garrison, T. Anderson, and W. Archer, “Critical thinking, cognitive presence and computer conferencing in distance education,” *American Journal of Distance Education*, vol. 15, pp. 7-23, 2001.

[44] D. L. Gladstein, “Groups in context: A model of task group effectiveness,” *Administrative Science Quarterly*, vol. 29, pp. 499–517, 1984.

[45] E. Goffman, *Interaction ritual: Essays on face-to-face behavior*. Garden City, NY: Anchor Books, 1967.

[46] D. S. Gouran, C. Brown, and D. R. Henry, “Behavioral correlates of perceptions of quality in decision-making discussions,” *Communication Monographs*, pp. 51-63, 1978.

[47] J. W. Graham, “Principled organizational dissent: A theoretical essay,” in *Research in Organizational Behavior* vol. 8, B. M. Staw and L. L. Cummings, Eds. Greenwich, CT: JAI Press, 1986, pp. 1-52.

[48] R. M. Grant, “Toward a knowledge-based theory of the firm,” *Strategic Management Journal*, vol. 17, pp. 109–122, 1996.

[49] R. Grishman and R. Kittredge, “Analyzing Language in Restricted Domains: Sublanguage Description and Processing.” Hillsdale, NJ: Lawrence Erlbaum, 1986.

[50] J. R. Hackman, “The design of work teams,” in *The Handbook of Organizational Behavior*, J. W. Lorsch, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1987, pp. 315–342.

[51] C. Hass, *Writing technology: Studies on the materiality of literacy.* . Manwah, N.J.: Erlbaum, 1996.

[52] R. Heckman and H. Annabi, “A content analytic comparison of FTF and ALN case-study discussions,” in *36th Annual Hawaii International Conference on System Sciences (HICSS'03)*. Big Island, Hawaii: IEEE Press, 2003.

[53] D. Hellriegel, S. E. Jackson, J. Slocum, and G. Staude, *Management*. Cape Town: OUP, 2001.

[54] J. D. Herbsleb, A. Mockus, T. A. Finholt, and R. E. Grinter, “An empirical study of global software development: Distance and speed,” presented at Proceedings of the International Conference on Software Engineering (ICSE 2001), Toronto, Canada, 2001.

[55] S. C. Herring, "Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives." Philadelphia: John Benjamins, 1996.

[56] S. R. Hiltz and M. Turoff, *The Network Nation*. Cambrige, MA: MIT Press, 1993.

[57] R. J. House and R. N. Aditya, "The social scientific study of leadership: Quo vadis? ," *Journal of Management*, vol. 23, pp. 409–473, 1997.

[58] J. Howison, M. Conklin, and K. Crowston, "FLOSSmole: A collaborative repository for FLOSS research data and analyses," *International Journal of Information Technology and Web Engineering*, In press.

[59] J. Howison, M. S. Conklin, and K. Crowston, "OSSmole: A collaborative repository for FLOSS research data and analyses," presented at 1st International Conference on Open Source Software, Genova, Italy, 2005.

[60] J. Howison and K. Crowston, "The perils and pitfalls of mining SourceForge," presented at Presentation at the Workshop on Mining Software Repositories, 26th International Conference on Software Engineering, Edinburgh, Scotland, 2004.

[61] S. L. Jarvenpaa and D. E. Leidner, "Communication and trust in global virtual teams," *Organization Science*, vol. 10, pp. 791–815, 1999.

[62] M. Konovsky and D. Organ, "Dispositional and contextual determinants of organizational citizenship behavior," *Journal of Organizational Behavior*, vol. 17, pp. 253-266, 1996.

[63] B. H. Kwasnik and K. Crowston, "A framework for creating a facetted classification for genres: Addressing issues of multidimensionality," in *Proceedings of the Hawai'i International Conference on System Science (HICSS)*. Big Island, Hawai'i, 2004.

[64] B. H. Kwasnik and K. Crowston, "Genres of digital documents: Introduction to the special issue," *Information, Technology & People*, vol. 18, pp. 76–88, 2005.

[65] G. F. Lanzara and M. Morner, "Making and sharing knowledge at electronic crossroads: the evolutionary ecology of open source," presented at 5th European Conference on Organizational Knowledge, Learning and Capabilities, Innsbruck, Austria, 2004.

[66] G. K. Lee and R. E. Cole, "From a firm-based to a community-based model of knowledge creation: The case of Linux kernel development," *Organization Science*, vol. 14, pp. 633–649, 2003.

[67] E. D. Liddy, "Breaking the Metadata Generation Bottleneck," in *Joint Conference on Digital Libraries*. Roanoke, VA, 2001.

[68] E. D. Liddy, "Data-Mining, MetaData, and Digital Libraries," in *DIMACS Workshop on Data Analysis and Digital Libraries*. Rutgers University, 2001.

[69] E. D. Liddy, "Automating and Evaluating Automatic Metadata Generation," in *Search Engine Conference*. Boston, MA, 2003.

[70] E. D. Liddy, "Automating and Evaluating Metadata Generation," in *InfoToday 2003 Conference*. New York, NY, 2003.

[71] E. D. Liddy, "Automating and Evaluating Metadata Generation," in *Libraries in the Digital Age Conference*. Dubrovnik, Croatia., 2003.

[72] E. D. Liddy, "Natural Language Processing for Text Extraction Applications. Keynote Speaker," presented at Thomson 8th Annual Text Summit, Minneapolis, MN, 2003.

[73] E. D. Liddy, "Generating & evaluating automatic metadata for educational resources," presented at European Conference on Digital Libraries, Vienna, 2005.

[74] E. D. Liddy, "MetaTest: A Tripartite Evaluation," presented at American Society for Information Science & Technology Conference, Charlotte, NC, 2005.

[75] E. D. Liddy and A. R. Diekema, "NLP for NSDL and All of NSF! Invited Talk," National Science Foundation, 2006.

[76] E. D. Liddy and C. Finneran, "Developing and Evaluating Metadata for Improved Information Access," in *NSF-NSDL Annual Meeting*. Washington, DC, 2003.

[77] E. D. Liddy and C. Finneran, "MetaTest: Three-Way Evaluation of Automatic Metadata Generation," in *Joint IMLS/NSDL Conference*, 2003.

[78] E. D. Liddy, G. Gay, S. Harwell, and T. Finneran, "A Modest (Metadata) Proposal," in *Joint Conference on Digital Libraries*. Portland, OR, 2002.

[79] E. D. Liddy, C. L. Jorgensen, E. E. Sibert, and E. S. Yu, "Sublanguage grammar in natural language processing," in *Proceedings of RIAO '91 Conference*. Barcelona, 1991.

[80] E. D. Liddy, C. L. Jorgensen, E. E. Sibert, and E. S. Yu, "A sublanguage approach to Natural Language Processing for an expert system," *Information Processing & Management*, vol. 29, pp. 633–645, 1993.

[81] M. Lipman, "Philosophy for Children," *Metaphilosophy*, vol. 7, pp. 17-39, 1976.

[82] L. L. Martins, L. L. Gilson, and M. T. Maynard, "Virtual teams: What do we know and where do we go from here?," *Journal of Management*, vol. 30, pp. 805-835, 2004.

[83] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis : An Expanded Sourcebook*, 2nd ed. Thousand Oaks: Sage Publications, 1994.

[84] D. A. Morand and R. J. Ocker, "Politeness theory and Computer-Mediated communication: A Sociolinguistic Approach to Analyzing Relational Messages. ," in *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003.

[85] M. D. Myers and K. Crowston, "Will real estate agents survive? The transformation of the real estate industry by information technology," *University of Auckland Business Review*, vol. 6, 2004.

[86] D. R. Newman, C. Johnson, C. Cochrane, and B. Webb, "An experiment in group learning technology: evaluating critical thinking in face-to-face and computer-supported seminars," *Interpersonal Computer and Technology Journal*, vol. 4, pp. 57-74, 1996.

[87] M. O'Leary and J. Cummings, "The Spatial, Temporal, and Configurational Characteristics of Geographic Dispersion in Teams," presented at Academy of Management Conference, Denver, CO, 2002.

[88] R. J. Ocker, "The Relationship between Interaction, Group Development, and Outcome: A Study of Virtual Communication," in *Proceedings of the Thirty-forth Annual Hawaii International Conference on System Science*, 2001.

[89] R. Ohlsson, "An early form of the community of inquiry: The study circle," *Thinking*, vol. 14, 1998.

[90] D. R. Olson, *The world on paper: The conceptual and cognitive implications of reading and writing* Cambridge and New York: Cambridge University Press, 1994.

[91] G. M. Olson and J. S. Olson, "Distance matters," *Human-Computer Interaction*, vol. 15, pp. 139–178, 2000.

[92] G. M. Olson, Teasley, S., Bietz, M.J., & Cogburn, "Collaboratories to support distributed science: The example of international HIV/AIDS research," presented at Annual research conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology, Port Elizabeth, South Africa, 2002.

[93] D. Organ, *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books, 1988.

[94] D. Organ, P. Podsakoff, and S. MacKenzie, *Organizational citizenship behavior: Its nature, antecedents, and consequences*. Thousand Oaks, CA: SAGE Publications, 2006.

[95] C. Pavitt, "Small Group Communication: A Theoretical Approach (3rd Ed)," vol. 2004, 1998.

[96] O. Rambow, L. Shrestha, J. Chen, and C. Laurdisen, "Summarizing Email Threads.," in *Proceedings of HLT-NAACL*, 2004, pp. 105–108.

[97] E. S. Raymond, "The cathedral and the bazaar," *First Monday*, vol. 3, 1998.

[98] M. Ridley, *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. New York: Viking, 1996.

[99] T. B. Roby, "The executive function in small groups," in *Leadership and Interpersonal Behavior*, L. Petrullo and B. M. Bass, Eds. New York: Holt, Rinehart and Wilson, 1961.

[100] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer, "Assessing social presence in asynchronous, text-based computer conferencing," *Journal of Distance Education*, vol. 14, pp. 51–70, 1999.

[101] N. Sager, C. Friedman, and M. S. Lyman, *Medical Language Processing: Computer Management of Narrative Data*. Reading, Mass: Addison-Wesley, 1987.

[102] S. Sawyer, K. Crowston, R. Wigand, and M. Allbritton, "The social embeddedness of transactions: Evidence from the residential real estate industry," *The Information Society*, vol. 19, pp. 135–154, 2003.

[103] S. Sawyer, R. Wigand, and K. Crowston, "Redefining access: Uses and roles of information and communication technologies in the US residential real estate industry from 1995 to 2005," *Journal of Information Technology*, vol. 20, pp. 213–233, 2005.

[104] W. C. Schutz, "The ego, RIFO theory, and the leader as completer," in *Leadership and Interpersonal Behavior*, L. Petrullo and B. M. Bass, Eds. New York: Holt, Rinehart and Wilson, 1961, pp. 48–65.

[105] A. Sharp, "What is a community of inquiry?," *Journal of Moral Education*, vol. 16, pp. 37–44, 1987.

[106] C. Smith, D. Organ, and J. Near, "Organizational citizenship behavior: Its nature and antecedents," *Journal of Applied Psychology*, pp. 653-663, 1983.

[107] A. Stein, J. Keiper, L. Bezerra, H. Brocks, and U. Thiel, "Collaborative research and documentation of European film history: The COLLATE Collaboratory," *Journal of Digital Information Management*, vol. 2, pp. 30–39, 2004.

[108] W. Turner, J.-P. Sansonnet, L. Gasser, and G. Ripoche, "Confidence-based organizational metrics," presented at Workshop on Distributed Collective Practice: Building new Directions for Infrastructural Studies, CSCW 2004, 2004.

[109] E. von Hippel, "Innovation by user communities: Learning from open-source software," *Sloan Management Review*, pp. 82–86, 2001.

[110] E. von Hippel and G. von Krogh, "Exploring the Open Source Software Phenomenon: Issues for Organization Science," Sloan School of Management, MIT, Cambridge, MA 2002.

[111] E. von Hippel and G. von Krogh, "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science," *Organization Science*, vol. 14, pp. 209–213, 2003.

[112] R. E. Walton and J. R. Hackman, "Groups under contrasting management strategies," in *Designing Effective Work Groups*, P. S. Goodman and Associates, Eds. San Francisco, CA: Jossey-Bass, 1986, pp. 168–201.

[113] P. Wayner, *Free For All*. New York: HarperCollins, 2000.

[114] E. Webb and K. E. Weick, "Unobtrusive measures in organizational theory: A reminder," *Administrative Science Quarterly*, vol. 24, pp. 650–659, 1979.

[115] E. M. White, "Assessing higher-order thinking and communication skills in college graduates through writing," *The Journal of General Education*, pp. 105-122, 1993

[116] L. Williams and S. Anderson, “Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors,” *Journal of Management*, pp. 601-617, 1991.

[117] W. A. Wulf, “Towards a National Collaboratory,” Report of Invitational Workshop at Rockefeller University 1989.

[118] W. A. Wulf, “National Collaboratories: Applying Information Technology for Scientific Research,” National Academy Press., Washington, DC 1993.

[119] O. Yilmazel, C. M. Finneran, and E. D. Liddy, “MetaExtract: An NLP System to Automatically Assign Metadata,” in *Proceedings of the 2004 Joint Conference on Digital Libraries*, 2004.

[120] G. Yukl, *Leadership In Organizations*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[121] G. Yukl, A. Gordon, and T. Taber, “A hierarchical taxonomy of leadership behavior: Integrating a half century of behavior research,” *Journal of Leadership & Organization Studies*, vol. 9, pp. 15-32, 2002.

[122] N. Zakaria and D. L. Cogburn, “Webs of culture: The influence of cultural context on decision-making in transnational NGO networks in WSIS,” presented at 2006 International Communication Association Annual Meeting, Dresden, Germany, 2006.

[123] I. Zigurs, “Leadership in virtual teams: Oxymoron or opportunity?,” *Organizational Dynamics*, vol. 31, pp. 339-351, 2003.

Budget Justification

A. Salaries and Wages – Senior Personnel

The PIs, Drs. Derrick Cogburn, Kevin Crowston, Robert Heckman and Elizabeth Liddy will work during the summer ($4,500 per PI per summer, approximately 0.4 summer month). Summers will be devoted to sample selection, detailed project design, integration of data analysis and publication of results. All PIs will devote 10% of effort during the academic year to project management and oversight (1/2 day / week, supported by Syracuse University). Dr. Crowston will be responsible for overall project direction and coordination, for assuring successful project completion, including submission of NSF progress reports, as required. The PIs will jointly be responsible for the review of the data and preparation of manuscripts for publication.

B. Salaries and Wages – Other Personnel

Approximately 3/4 of the direct funding is requested to support PhD student tuition and stipends. Stipends are requested for two Ph.D. students, 50% academic year and 100% summer effort, for a total of 2200 hours/year (4400 hours in two years). The graduate students will support the principal investigators in sample section and will have primary responsibility for data collection and analysis, under the oversight of the PIs.

C. Fringe Benefits

Fringe Benefits are calculated as direct costs in accordance with Syracuse University's indirect cost rate agreement (Department of Health and Human Services, 17.5% for faculty during the summer, 14.7% for graduate students).

E2. Travel:

Travel support is requested for students and PIs to disseminate results at academic conferences (one trip each, approximately $1300/trip).

G. Other Direct Costs

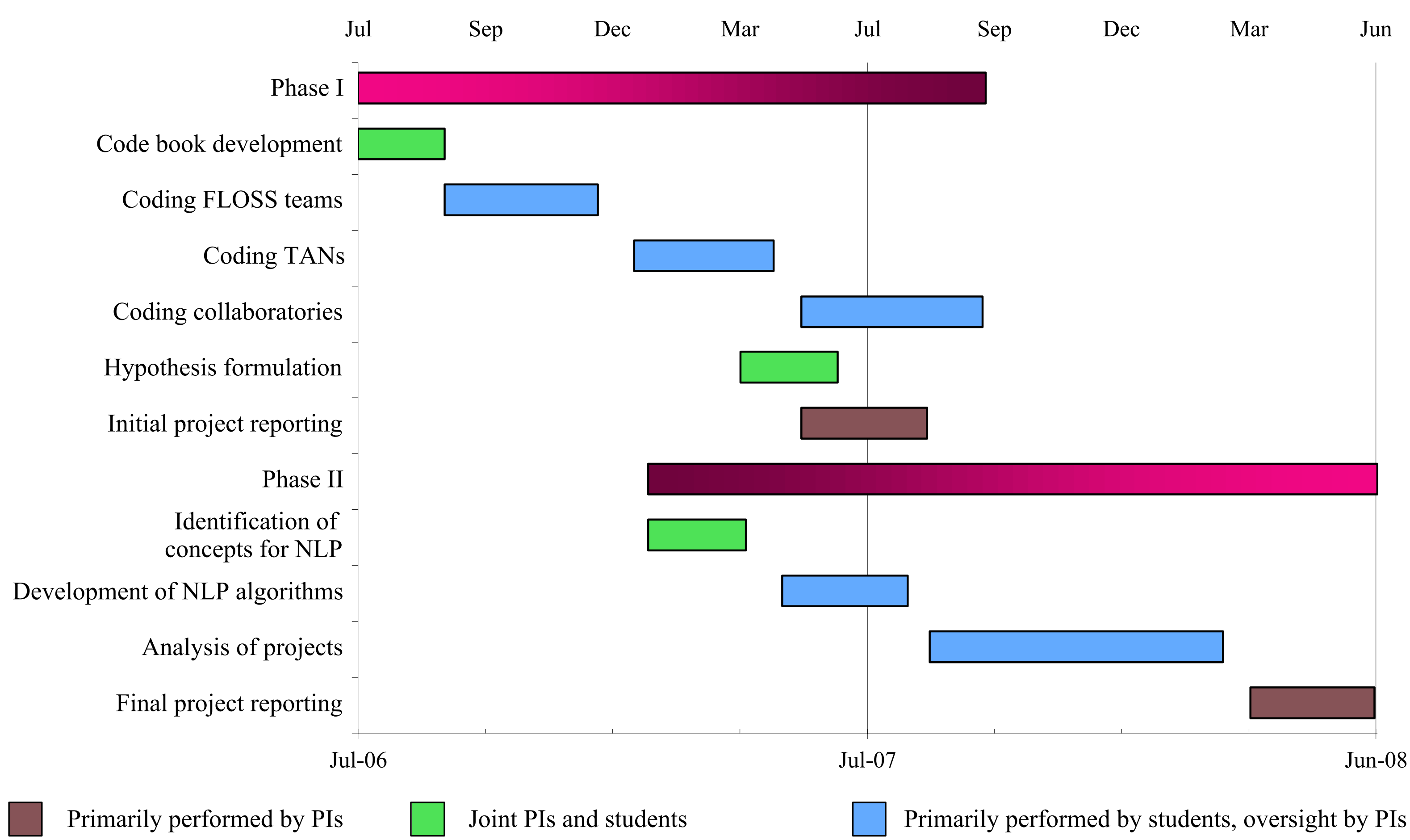
6. Other:

A total of $47,016 is requested for partial support of tuition for two graduate students (12 credit hours per year, per student at $941/credit for Year 1 and $1,018/credit for Year 2). Any additional required tuition will be supported by Syracuse University.

I. Indirect Costs

Indirect Costs are calculated in accordance with Syracuse University's federally negotiated indirect cost rate agreement (Department of Health and Human Services), which is currently 47.5% of modified total direct costs (MTDC).

The following proposed timeline for the project indicates how the requested resources will be applied.



Facilities, equipment and other resources

Syracuse University is one of the largest and most comprehensive independent universities in the United States. Founded in 1870, Syracuse offers excellent facilities, equipment and other resources for research and study in many academic and professional disciplines.

The **School of Information Studies** is a leading center for innovative programs in information policy, information behavior, information management, information systems, information technology and information services. Its approach stands out from other institutions that offer computer science, management, information science and related programs in that our focus is on users and user information needs as a starting point for integrating information and information technology into organizations. The faculty of the School crosses disciplinary boundaries to integrate the common elements of information management in business, government, education, and nonprofit settings, including the relationship of information and knowledge, electronic and traditional libraries, information systems and technology, information resources management, information policy and services, and the study of information users.

The School has seven active research centers, of which one, the **Center for Natural Language Processing**, will be central in this research. CNLP advances the development of human-like language understanding software capabilities for government, commercial, and consumer applications. It is situated in its own lab facilities in Hinds Hall at Syracuse University. The Center for Natural Language Processing has five servers, and twenty-one computers. In addition to its own lab space and equipment, the Center has access to the meeting rooms, labs, and classroom space of the School of Information Studies. The Center also has access to technical and administrative resources within the greater University.

The Center has been successful at attracting top student talent for its many Research Assistantships, including two PhD students who have won the prestigious ISI Doctoral Dissertation Proposal Award and the ProQuest Doctoral Dissertation Award presented by the American Society for Information Science and Technology

The School's other research centers are:

- Center for Digital Commerce. Conducts research and provides strategic analyses in all areas of digital and electronic commerce.
- Center for Emerging Network Technologies. Performs hands-on testing and provide industry analysis of products and services in emerging technology markets.
- The Convergence Center. Supports research on and experimentation with media convergence to understand the future of digital media and to engage students and faculty in the process of defining and shaping that future.
- The Systems Assurance Institute, a collaboration among Engineering and Computer Science, Information Studies, the Newhouse School of Public Communications and the Maxwell School of Citizenship and Public Affairs. Advances the understanding and state-of-the-practice of systems assurance.
- The Center for Digital Literacy. Supports collaborative research and development projects related to understanding the impact of information, technology and media literacies on children and adults in today's technology-intensive society.
- The Information Institute of Syracuse (IIS) (http://iis.syr.edu/). The umbrella organization for a number of highly visible and widely successful digital education information services to improve learning and teaching in the U.S. and throughout the world.

The School of Information Studies space plan includes providing (1) a space for a community of learning, research, and education for students and faculty; (2) space that supports economic development and growth in Central New York: (3) space that supports research, development and economic growth through the School's research centers; (4) common spaces that are inviting to students and visitors; (5) space that supports communication and connections between floors to preserve the strong feelings among students, faculty, and staff of being on the IST team; (6) a building that supports state of the art technology including broadband and wireless in offices, classrooms and centers; (7) space with the flexibility to change to meet the needs of a changing networked economy, changing technology, research, and faculty and student needs; (8) classroom space that supports student access to technology and/or classroom discussions in a room such as a case management classroom; (9) sufficient conference and meeting room space for a school enriched by its faculty and staff commitment to team meetings, service, and collaborative research; and (10) space that supports a collaborative learning environment for students.

SU's library system serves the information and research needs of the academic community. The collections exceed 2.6 million volumes, 11,330 serials and periodicals, and 3.4 million microforms, located in several libraries on campus. Library services include information and reference, online database searching, access to bibliographic and other data on CD-ROM and interlibrary loan.

Computing Services helps researchers, faculty and students use computing by providing personal computers, mainframe computers, data communication networks, software, training and advice. Most equipment and services are available without a direct charge.